

COMPLEMENTARITY AND IDENTIFICATION

TATE TWINAM

*Department of Economics
University of Pittsburgh*

ABSTRACT. This paper examines the identification power of assumptions that formalize the notion of complementarity in the context of a nonparametric bounds analysis of treatment response. I extend the literature on partial identification via shape restrictions by exploiting cross-dimensional restrictions on treatment response when treatments are multidimensional; the assumption of *supermodularity* can strengthen bounds on average treatment effects in studies of policy complementarity. I combine this restriction with a statistical independence assumption to derive improved bounds on treatment effect distributions, aiding in the evaluation of complex randomized controlled trials. I show how complementarities arising from treatment effect heterogeneity among subpopulations can be incorporated through *super-modular instrumental variables* to strengthen identification of treatment effects in studies with one or multiple treatments. I use these results to examine the long-run effects of zoning on the evolution of land use patterns.

1. INTRODUCTION

Complementarities arise naturally in many economic problems, often manifesting as policy interactions or treatment effect heterogeneity among observed subgroups of a population. This paper examines how assumptions that formalize the notion of complementarity can aid in the identification of treatment effects. I employ a nonparametric bounds approach, where identification is driven by qualitative restrictions rooted in economic theory or empirical evidence rather than strong functional form or unconfoundedness assumptions. This approach will yield interval estimates of parameters of interest; however, informative bounds are often preferable to precise (but wrong) estimates obtained under incorrect assumptions. Partial identification tools have been fruitfully applied to a wide range of empirical problems.¹

E-mail address: tat47@pitt.edu.

Date: September 5, 2014.

JEL classification. C31, C36.

Key words and phrases. Partial identification, nonparametric bounds, supermodularity, instrumental variables, program evaluation.

¹Examples of applied partial identification studies include: Giustinelli (2011) and Tsunao and Usui (2014) on the returns to education, Kreider and Pepper (2007) on disability and employment,

In particular, I explore the identification power yielded by assuming that individual treatment response functions exhibit *supermodularity* when treatments are multidimensional. This assumption allows one to construct more informative bounds in studies of policy complementarity, which are typically stymied by the absence of pseudo-experimental variation in the assignment of multiple treatments. I also show how complementarities arising from interactions between treatment effects and observable covariates can be formalized as *supermodular instrumental variables* to improve bounds on average treatment effects. This novel instrumental variable approach is broadly applicable to studies with one or multiple treatments. Complementarity is frequently invoked in economics, but studies of its identification power have been limited to very specific contexts. This paper develops general results applicable to program evaluation in a wide range of empirical situations. I illustrate the use of my results in an empirical application on the long-run effects of zoning on land use patterns.

Typically, empirical studies seek to estimate the effect of a single treatment on one or more outcome variables. However, the effect of a treatment may vary substantially with the value of other (endogenously-determined) treatment variables. When policymakers have multiple tools at their disposal, understanding how different policies enhance or offset each other is crucial. If the positive impact of some policy intervention is substantially larger when combined with a second (costly) intervention, a measure of the magnitude of this difference is necessary for a proper cost-benefit analysis. The supermodularity and submodularity assumptions I propose can aid in quantifying how policy impacts differ with the associated policy environment.²

For example, unemployment relief is a multidimensional policy, involving a choice of both potential benefit duration and the wage replacement rate. Lalive, Van Ours and Zweimüller

Bhattacharya, Shaikh and Vytlačil (2008, 2012) on the mortality effects of Swan-Ganz catheterization, Kreider and Hill (2009) on the effect of universal health insurance on medical expenditures, Pepper (2000) on the intergenerational transmission of welfare receipt, Manski and Nagin (1998) on sentencing and recidivism, and Gundersen, Kreider and Pepper (2012) on the health effects of the National School Lunch Program.

²As more treatments are considered, the data are necessarily less informative about each individual treatment. Thus, the researcher faces a trade-off where richer treatment spaces allow for more interesting questions but generally lead to less precise answers.

(2006) show both theoretically and empirically that these two dimensions are complementary, with simultaneous increases in both the replacement rate and potential benefit duration leading to an increase in unemployment duration substantially larger than the sum of the effects measured individually for particular subgroups. The Lalive et al. study exploits variation in both dimensions of unemployment relief that has the characteristics of a natural experiment, but such opportunities are very rare. Pseudo-experimental variation along multiple policy dimensions is far less common than similar variation in individual policies. This has arguably led to the overwhelming focus on the effects of policies in isolation. The partial identification tools developed here, which are applicable in the absence of any unusual pseudo-experimental policy variation, should enhance the ability of researchers to measure treatment effect heterogeneity due to policy complementarities in a wide range of contexts.³ I illustrate the use of the shape restrictions developed here in a study of how the long-run effect of commercial zoning on land use patterns varies with different restrictions on building density.

Relatedly, responses to a treatment may differ among subpopulations defined by observable covariates. Many recent experimental studies have discussed the importance of treatment effect heterogeneity between subgroups (Bitler, Gelbach and Hoynes 2006, 2008, 2014, Djebbari and Smith 2008, Feller and Holmes 2009). I show how qualitative information about such treatment effect heterogeneity leads naturally to supermodular instrumental variables, which can help narrow the bounds on average treatment effects in the same manner as a traditional instrumental variable or a monotone instrumental variable.⁴ Supermodular instrumental variables can be applied in the case of a single treatment or multiple treatments, making them a potentially valuable addition to the range of identifying assumptions available to applied researchers. I demonstrate their utility in the empirical illustration in section 6.

³The sensitivity of effects to the surrounding policy environment may partly explain the wide variation in estimates of treatment effects for similar policies in different contexts found in many literatures; see, for example, the discussion in Lalive et al. (2006) on the effects of unemployment benefit policies on re-employment rates. See also Gelman (2013) for a related discussion.

⁴See Manski and Pepper (2000) and section 4.

While the bulk of the paper focuses on identification using non-experimental data, the assumptions developed in this paper can be applied in the evaluation of complex randomized controlled trials (RCTs) involving multiple treatments. In a discussion of program evaluation, Heckman, Smith and Clements (1997) note that, even in an RCT, many parameters of interest are not point-identified, such as the proportion of the population receiving a treatment who benefit from the treatment. Heckman et al. observe that classical probability inequalities like the Fréchet-Hoeffding copula bounds are not very informative. The structural supermodularity and submodularity assumptions I introduce have implications for the entire distribution of treatment effects, so they can be used to obtain stronger bounds. Since average treatment effects are identified in this context, the supermodularity or submodularity of average effects can be established, and this can be used to provide some justification for the stronger structural assumptions. Similarly, the validity of supermodular instrumental variable assumptions can be established and used to justify stronger *quantile supermodular instrumental variable* assumptions, which can also be applied in the case of a single treatment.

The literature on partial identification is extensive.⁵ Many of the contributions of Charles Manski and coauthors are relevant to the results developed below; I review them as appropriate. The literature on complementarity and identification is relatively small. Molinari and Rosen (2008) connect supermodularity to identification in the context of game estimation. They show that the approach of Aradillas-Lopez and Tamer (2008) applies to games with supermodular payoff functions. Eeckhout and Kircher (2011) find that they cannot identify (using wage data alone) whether or not the technology of a firm is supermodular, i.e., whether or not more productive workers sort towards more productive jobs. Graham, Imbens and Ridder (2014) analyze how reallocations of indivisible heterogeneous inputs across production units (leaving a potentially complementary input fixed) may affect average output. They discuss identification and estimation of the effects of a variety of correlated matching rules. Lazzati (2014) uses monotone comparative statics to partially identify treatment response in the

⁵See Manski (2003) for a comprehensive overview.

presence of endogenous social interactions. Supermodularity arises naturally in this context when individual outcomes are increasing with the outcomes of others. The shape restrictions I propose have been used in the context of estimation to improve efficiency; Beresteanu (2005, 2007) considers the efficiency gains from imposing a variety of restrictions, including supermodularity and submodularity.

The remainder of the paper is organized as follows. In section 2, I outline the formal setup used throughout the paper. In section 3, I present novel shape restrictions and the resulting bounds on average treatment effects. In section 4, I discuss instrumental variable assumptions and derive bounds on average potential outcomes and average treatment effects. In section 5, I combine shape restrictions and instrumental variables with statistical independence assumptions to derive bounds on cumulative distribution functions of treatment effects. I conclude with an empirical illustration on the long-run effects of zoning on land use patterns in section 6.

2. NOTATION AND SETUP

Individuals are drawn from a population I . The set I , the Borel σ -algebra of subsets of I denoted by \mathcal{I} , and the probability measure P together form a probability space (I, \mathcal{I}, P) . Every individual $i \in I$ is associated with a vector of covariates $x^i \in X$ and a vector of realized treatments $z^i \in T$, where T is the treatment set.⁶ Since I focus on the identification of treatment effects in the presence of multiple treatments, I will discuss in detail the structure I adopt for the treatment space.

Definition. A nonempty partially ordered set V is a *lattice* if, for any $v, v' \in V$,

- V contains the join (least upper bound) of v and v' , denoted by $v \vee v'$, and
- V contains the meet (greatest lower bound) of v and v' , denoted by $v \wedge v'$.

Examples of lattices include \mathbb{R}^2 , $\mathbb{Z} \times \mathbb{R}$, and $\{0, 1\}^n$ for $n \in \mathbb{N}$. The meet and join operations depend on the particular order imposed on the lattice; for example, the join of $(2, 0)$ and $(1, 1)$ in \mathbb{R}^2 is equal to $(2, 1)$ under the product order and $(2, 0)$ under the lexicographic

⁶I use superscripts to refer to individuals and reserve subscripts to denote vector components.

order. An element v of a lattice V is the top (bottom) of V if $v' \leq v$ ($v \leq v'$) for all $v' \in V$; if v is not the top or bottom, it is in the interior. If the top (or bottom) of a lattice exists, it is unique.

Definition. For a lattice V , a nonempty subset $U \subseteq V$ is a *sublattice* of V if, for any $u, u' \in U$, U contains the meet and join of u and u' in V .

Sublattices will be useful when I consider assumptions that do not hold globally on T . The following assumption, which I maintain throughout the paper, describes the structure imposed on the treatment space:

Assumption. The treatment space T is such that

- $T \subseteq \mathbb{R}^L$ with $L \in \mathbb{N}$,
- T is partially ordered under the product order, and
- T is a nonempty lattice.

The product order on T implies that $t \leq t'$ iff $t_l \leq_l t'_l$ for each l . If $t, t' \in T$ are incomparable, i.e., $t_l < t'_l$ and $t'_{l'} < t_{l'}$ for some l, l' , I write $t \parallel t'$. The advantage of the lattice assumption is the notational clarity it provides when I employ supermodularity and submodularity to formalize how the marginal effect on the response variable of changes in some dimensions of the treatment depend on the values of other dimensions of the treatment.

This specification is flexible enough to allow for a wide variety of treatment types. In this paper, I restrict attention to discrete treatments, as these are most commonly encountered in practice. Dimensions of the treatment may be binary or multi-valued (Cattaneo 2010). Most of the results extend straightforwardly to the case of continuous treatments. In practice, however, the application to continuous treatments is hampered by the fact that, as the number of treatments increases, the data alone are increasingly uninformative about the effect of each individual treatment. The relationship between the complexity of the treatment set and the amount that can be learned from the data is an issue I will discuss further in the next section.

Every individual i is associated with a (measurable) response function $y^i(\cdot) : T \rightarrow Y \in \mathbb{R}$ mapping treatments into outcomes $y^i(t)$.⁷ $z^i \in T$ is the treatment that i actually receives, so $y^i(z^i)$ is individual i 's realized outcome, $\{y^i(t)\}_{t \neq z^i}$ are individual i 's counterfactual outcomes, and $\{y^i(t)\}_{t \in T}$ are individual i 's potential outcomes. Throughout the paper, I maintain the stable unit treatment value assumption,⁸ which says that individuals' potential outcomes $\{y(t)\}_{t \in T}$ do not depend on other individuals' realized treatments (Rubin 1978).

3. SHAPE RESTRICTIONS

In this section, I explore the identifying power of shape restrictions that formalize complementarity and substitutability, with an emphasis on the identification of average treatment effects. I review shape restrictions proposed in the previous literature before moving on to the novel restrictions I propose. Using these assumptions, I derive bounds on average treatment effects for both simple and complex treatment spaces.

Throughout, I assume that there exist $\underline{K}, \overline{K} \in \mathbb{R}$ such that $\underline{K} \leq y(t) \leq \overline{K}$ for all t ; these are global bounds on response functions. As Manski (1990) observes, this is not as restrictive as it seems; for example, if y is a probability, it is naturally bounded between zero and one. In the absence of these global bounds, the results below will generally be uninformative. All well-defined expectations are assumed to exist; if an expectation $E[y(t) \mid z = t']$ is ill-defined because the event $z = t'$ is off the support of z , I establish the convention that $E[y(t) \mid z = t']P(z = t') \equiv 0$.

Manski (1989) introduced the no-assumption bounds on $E[y(t)]$. The no-assumption upper bound is the average of $E[y(t) \mid z = t]$ and the global upper bound \overline{K} , weighted respectively by $P(z = t)$ and $P(z \neq t)$; likewise for the lower bound. Since they are typically wide, research has focused on other credible assumptions that yield additional identifying power.

⁷I suppress i when referring to arbitrary response functions, covariates, or realized treatments.

⁸This assumption is alternatively referred to as noninterference by Cox (1958) and individualistic treatment response by Manski (2013).

Manski (1997) studied the identification power of assumptions on the shape of individual response functions; in particular, he considered restricting response functions to be monotone, semi-monotone, or concave-monotone. I reproduce the semi-monotone treatment response assumption here, in my notation:

Assumption SMTR (Semi-monotone treatment response). Response functions exhibit *semi-monotone treatment response* on $S \subseteq T$ if, for all $t, t' \in S$,

$$t \leq t' \implies y(t) \leq y(t')$$

If S is a chain,⁹ then this assumption is referred to as *monotone treatment response* (MTR).

Manski (1997) uses this assumption to derive bounds on numerous quantities, including average and quantile treatment effects. He motivated MTR by considering traditional demand analysis, where researchers often make strong parametric assumptions but do not exploit the less-controversial assumption that demand curves are downward sloping. SMTR removes the need for a totally ordered T . It has the same identification power regardless of whether $T \subseteq \mathbb{R}$ or $T \subseteq \mathbb{R}^L$ for $L > 1$, except that in the latter case, it is possible that $t \parallel t'$. Bhattacharya et al. (2008) derives bounds using SMTR without assuming a particular direction of monotonicity. Tsunao and Usui (2014) study the identification power of concave-monotone treatment response combined with monotone treatment selection (discussed in section 4).

MTR and SMTR are within-dimension restrictions on the response functions. Additional identification power can be obtained from cross-dimension restrictions, where the marginal effect of a change in some dimensions of the treatment variable depends on the values of the other dimensions:

Assumption SPM (Supermodularity). Response functions are *supermodular* on a sublattice $S \subseteq T$ if, for all $t, t' \in S$,

$$(3.1) \quad y(t') + y(t) \leq y(t \vee t') + y(t \wedge t')$$

⁹A subset S of a partially ordered set T is a *chain* if it is totally ordered under the inherited order.

They are *strictly supermodular* when the inequality is strict.

Assumption SBM (Submodularity). Response functions are *submodular* on a sublattice $S \subseteq T$ if, for all $t, t' \in S$,

$$(3.2) \quad y(t') + y(t) \geq y(t \vee t') + y(t \wedge t')$$

They are *strictly submodular* when the inequality is strict.

SPM is a formalization of the notion of complementarity. If two dimensions of a treatment, say t_1 and t_2 , are complementary, then the magnitude of the change in the response variable due to an increase in t_1 is increasing with t_2 . Thus, the two dimensions of the treatment act to amplify each others marginal effects. In the case of a linear model

$$(3.3) \quad y = \alpha + \beta t_1 + \delta t_2 + \gamma t_1 t_2$$

supermodularity is equivalent to the sign restriction $\gamma > 0$. SBM is a formalization of substitutability, the case where elements of the treatment may mitigate each others effects. Returning to (3.3), submodularity is equivalent to the sign restriction $\gamma < 0$. If both supermodularity and submodularity hold, response functions are said to be *modular*. Since these assumptions can be applied on sublattices of T , it is possible to allow some dimensions of a treatment to be complements while those same dimensions are substitutes with other dimensions.

As I discussed in the introduction, Lalive et al. (2006) study the Austrian labor market and find that the two dimensions of unemployment relief, potential benefit duration and the wage replacement rate, are complementary (strongly for some groups, weakly for others). This finding could motivate the use of SPM in assessments of these and similar policies in other contexts where the pseudo-random variation they exploit is absent.

Neumark and Wascher (2011) provide another example of policy complementarity in a study on the interaction between the Earned Income Tax Credit (EITC) and the minimum wage. They find that a higher minimum wage enhances the positive effect of the EITC on

the labor supply of single mothers; they find the opposite effect for childless individuals, suggesting a crowding-out effect. These findings suggest that assumptions SPM and SBM, respectively for each subgroup, could be applied in other studies on how the effect of minimum wage changes are influenced by the EITC or similar programs.

Another naturally multidimensional policy is zoning. Zoning laws typically regulate many aspects of the built environment; most broadly, they regulate both what types of uses are allowed (commercial, industrial, etc.) and how densely land can be developed (lot coverage of buildings, maximum height, etc.). The effects of specific zoning policies vary widely with the overall policy bundle. Shertzer, Twinam and Walsh (2014) study the long-run impact of the initial zoning of Chicago on a variety of modern land use outcomes. The long-run impact of historical commercial zoning on present-day commercial land use turns out to hinge critically on the associated density restrictions; commercial zoning has a substantially larger effect when paired with low-density zoning. This motivates the assumption of SBM in the empirical application in section 6.

Since assumptions SPM and SBM can be applied on sublattices of T , it is possible to allow some dimensions of a treatment to be complements while those same dimensions are substitutes with other dimensions. For example, consider $T = \{0, 1\}^3$ under the product order. Assumption SPM on the two sublattices

$$S_1 = \{(1, 1, 0), (1, 0, 0), (0, 1, 0), (0, 0, 0)\}$$

$$S_2 = \{(1, 1, 1), (1, 0, 1), (0, 1, 1), (0, 0, 1)\}$$

combined with assumption SBM on the five sublattices

$$S_3 = \{(1, 0, 0), (1, 0, 1), (0, 0, 0), (0, 0, 1)\}$$

$$S_4 = \{(1, 1, 0), (1, 1, 1), (0, 1, 0), (0, 1, 1)\}$$

$$S_5 = \{(0, 1, 0), (0, 1, 1), (0, 0, 0), (0, 0, 1)\}$$

$$S_6 = \{(1, 1, 0), (1, 1, 1), (1, 0, 0), (1, 0, 1)\}$$

$$S_7 = \{(1, 1, 0), (1, 1, 1), (0, 0, 0), (0, 0, 1)\}$$

yields complementarity between the first two dimensions of the treatment but substitutability between the first two (individually and jointly) and the third.

An inspection of the no-assumption bounds reveals that the amount one can learn about $E[y(t)]$ or $E[y(t) - y(t')]$ from the data alone depends on $P(z = t)$ and, in the latter case, $P(z = t')$. If $P(z = t)$ is small, the data are practically uninformative about $E[y(t)]$.¹⁰ Thus, the researcher faces a trade-off where richer treatment spaces (which entail a larger number of treatments) allow for more interesting questions but generally lead to less precise answers. Adding “nuisance” dimensions to the treatment space that allow for the application of additional SPM or SBM assumptions will generally not aid in the identification of treatment effects of interest.

In propositions 1 and 2, I show how SPM and SBM can be used to compute bounds on the expectations of average treatment effects. In general, these bounds will improve upon the no-assumption bounds in the case of multidimensional treatments; with only a single treatment, SPM and SBM have no identifying power. The simplest nontrivial lattice treatment space is $T = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$, which corresponds to a two-dimensional binary treatment. The following result shows the implications of supermodularity for identification on this simple treatment space:

Proposition 1. *Assume that $T = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. Assume that SPM holds on T . Then, the bounds*

$$\begin{aligned} & E[y(1, 0) | z = (1, 0)] P(z = (1, 0)) + \underline{K} P(z \neq (1, 0)) \\ & - E[y(0, 0) | z = (0, 0)] P(z = (0, 0)) - \overline{K} P(z \neq (0, 0)) \\ (3.4) \quad & \leq E[y(1, 0) - y(0, 0)] \leq \end{aligned}$$

$$E[y(1, 1) | z = (1, 1)] P(z = (1, 1)) + E[y(1, 0) | z = (1, 0)] P(z = (1, 0))$$

¹⁰In the case where one or more of the dimensions of the treatment are continuous, the data are necessarily uninformative about almost all of the treatments. This motivates my restriction to discretely-valued treatments.

$$\begin{aligned}
& +\overline{K}P(z \in \{(0,0), (0,1)\}) - E[y(0,1) | z = (0,1)]P(z = (0,1)) \\
& - E[y(0,0) | z = (0,0)]P(z = (0,0)) - \underline{K}P(z \in \{(1,0), (1,1)\})
\end{aligned}$$

and

$$\begin{aligned}
& E[y(1,1) | z = (1,1)]P(z = (1,1)) + E[y(1,0) | z = (1,0)]P(z = (1,0)) \\
& +\underline{K}P(z \in \{(0,0), (0,1)\}) - E[y(0,1) | z = (0,1)]P(z = (0,1)) \\
& - E[y(0,0) | z = (0,0)]P(z = (0,0)) - \overline{K}P(z \in \{(1,0), (1,1)\}) \\
(3.5) \quad & \leq E[y(1,1) - y(0,1)] \leq
\end{aligned}$$

$$\begin{aligned}
& E[y(1,1) | z = (1,1)]P(z = (1,1)) + \overline{K}P(z \neq (1,1)) \\
& - E[y(0,1) | z = (0,1)]P(z = (0,1)) - \underline{K}P(z \neq (0,1))
\end{aligned}$$

are sharp.¹¹ The no-assumption bounds remain sharp for $E[y(1,1) - y(0,0)]$ and each average potential outcome $E[y(\cdot)]$ defined on T .

Proof of proposition 1. First, I show that SPM does not improve upon the no-assumption bounds on potential outcomes. SPM implies that

$$y^i(1,0) + y^i(0,1) \leq y^i(1,1) + y^i(0,0)$$

For each i , exactly one of these outcomes is observed. The unobserved terms may take any value in $[\underline{K}, \overline{K}]$. When $z^i \neq (1,0)$, there are three cases to consider. If $z^i = (1,1)$, then SPM implies

$$y^i(1,0) \leq \overline{K} \leq y^i(1,1) + \overline{K} - \underline{K}$$

If $z^i = (0,1)$, then

$$y^i(1,0) \leq \overline{K} \leq \overline{K} + \overline{K} - y^i(0,1)$$

¹¹There is no guarantee that these bounds will be nonempty; if an assumption implies that the bounds on the parameter of interest are empty, the assumption is falsified by the data. This caveat applies to all the results that follow.

If $z^i = (0, 0)$, then

$$y^i(1, 0) \leq \overline{K} \leq \overline{K} + y^i(0, 0) - \underline{K}$$

Thus, it follows that

$$y^i(1, 0) \in \begin{cases} \{y^i(1, 0)\} & \text{if } z^i = (1, 0) \\ [\underline{K}, \overline{K}] & \text{if } z^i \in \{(0, 0), (0, 1), (1, 1)\} \end{cases}$$

Taking expectations yields the no-assumption bounds. A similar argument applies to the other elements of T .

The SPM inequality does permit strengthened identification results for treatment effects. In the no-assumption case, if $z^i = (1, 1)$ or $z^i = (0, 1)$, then $y^i(1, 0) - y^i(0, 0) \in [\underline{K} - \overline{K}, \overline{K} - \underline{K}]$. Under SPM, the fact that we observe one of $\{y^i(1, 1), y^i(0, 1)\}$ allows us to further reduce this upper bound. Sharp bounds for the treatment effects $y^i(1, 0) - y^i(0, 0)$, $y^i(1, 1) - y^i(0, 1)$, and $y^i(1, 1) - y^i(0, 0)$ are given below.

$$(3.6) \quad y^i(1, 0) - y^i(0, 0) \in \begin{cases} [\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)] & \text{if } z^i = (0, 0) \\ [y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i = (1, 0) \\ [\underline{K} - \overline{K}, \overline{K} - y^i(z^i)] & \text{if } z^i = (0, 1) \\ [\underline{K} - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i = (1, 1) \end{cases}$$

$$(3.7) \quad y^i(1, 1) - y^i(0, 1) \in \begin{cases} [\underline{K} - y^i(z^i), \overline{K} - \underline{K}] & \text{if } z^i = (0, 0) \\ [y^i(z^i) - \overline{K}, \overline{K} - \underline{K}] & \text{if } z^i = (1, 0) \\ [\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)] & \text{if } z^i = (0, 1) \\ [y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i = (1, 1) \end{cases}$$

$$(3.8) \quad y^i(1, 1) - y^i(0, 0) \in \begin{cases} [\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)] & \text{if } z^i = (0, 0) \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z^i = (1, 0) \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z^i = (0, 1) \\ [y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i = (1, 1) \end{cases}$$

Taking expectations in equations (3.6) and (3.7) yields the bounds in (3.4) and (3.5), respectively. Equation (3.8) shows that the no-assumption bounds remain sharp for $E[y(1, 1) - y(0, 0)]$. \square

In proposition 1, assumption SPM improves the upper bound on $E[y(1, 0) - y(0, 0)]$ and the lower bound on $E[y(1, 1) - y(0, 1)]$ by establishing a monotonicity relationship between the two treatment effects. This monotonicity relationship implies that this assumption will allow improvements in the bounds on one treatment effect (due to the imposition of other assumptions) to further improve the bounds on other treatment effects. Because of this, bounds computed jointly under SPM and other assumptions like monotone or supermodular instrumental variables¹² will generally be strictly contained within the intersection of the bounds computed under these assumptions separately. The empirical application in section 6 illustrates this phenomenon. Thus, while SPM may have substantial identifying power on its own, it may yield even more identifying power when combined with other assumptions.

In the special case where y is bounded between zero and one, SPM can establish that $E[y(1, 0) - y(0, 0)] \in [-1, 0]$ or $E[y(1, 1) - y(0, 1)] \in [0, 1]$ if the observed expectations in (3.4) and (3.5) take certain boundary values. In general, however, SPM is not sufficient to identify the sign of a treatment effect in the absence of other assumptions.

Sharp bounds can be derived on general treatment spaces using the same approach, as I show in proposition 2:

¹²See section 4.

Proposition 2. Let $\{S_\gamma\}_{\gamma \in \Gamma}$ be the collection of all sublattices of T such that, for every $\gamma \in \Gamma$, S_γ is not a chain and $|S_\gamma| = 4$. Define $\Gamma^{SPM} \subseteq \Gamma$ to be the set of γ such that SPM holds on S_γ and SBM does not hold on S_γ iff $\gamma \in \Gamma^{SPM}$; likewise, define $\Gamma^{SBM} \subseteq \Gamma$ to be the set of γ such that SBM holds on S_γ and SPM does not hold on S_γ iff $\gamma \in \Gamma^{SBM}$. Define $\Gamma^{MOD} \subseteq \Gamma$ to be the set of γ such that both SPM and SBM hold on S_γ iff $\gamma \in \Gamma^{MOD}$. Let $\Gamma_{t,t'}^{SPM} \subseteq \Gamma^{SPM}$ be the set of γ such that $t, t' \in S_\gamma$ and $\gamma \in \Gamma^{SPM}$; likewise for $\Gamma_{t,t'}^{SBM}$ and $\Gamma_{t,t'}^{MOD}$. Then, for $t' < t$,

$$\begin{aligned}
& [\mathbb{E}[y(t) \mid z = t] - \overline{K}] P(z = t) + [\underline{K} - \mathbb{E}[y(t') \mid z = t']] P(z = t') \\
& + \sum_{t'' \in \Lambda_1 \cup \Lambda_3} [\mathbb{E}[y(t'') \mid z = t''] - \overline{K}] P(z = t'') \\
& + \sum_{t'' \in \Lambda_4 \cup \Lambda_5} [\underline{K} - \mathbb{E}[y(t'') \mid z = t'']] P(z = t'') \\
& + \sum_{t'' \in \Lambda_2 \cup \Lambda_6 \cup \Lambda_7} [\underline{K} - \overline{K}] P(z = t'') \\
(3.9) \quad & \leq \mathbb{E}[y(t) - y(t')] \leq
\end{aligned}$$

$$\begin{aligned}
& [\overline{K} - \mathbb{E}[y(t') \mid z = t']] P(z = t') + [\mathbb{E}[y(t) \mid z = t] - \underline{K}] P(z = t) \\
& + \sum_{t'' \in \Lambda_1 \cup \Lambda_2} [\mathbb{E}[y(t'') \mid z = t''] - \underline{K}] P(z = t'') \\
& + \sum_{t'' \in \Lambda_4 \cup \Lambda_6} [\overline{K} - \mathbb{E}[y(t'') \mid z = t'']] P(z = t'') \\
& + \sum_{t'' \in \Lambda_3 \cup \Lambda_5 \cup \Lambda_7} [\overline{K} - \underline{K}] P(z = t'')
\end{aligned}$$

where $\Lambda_1, \dots, \Lambda_7$ are defined in (3.11). These bounds are sharp.

Proof of proposition 2. By the Law of Iterated Expectations,

$$(3.10) \quad \mathbb{E}[y(t) - y(t')] = \sum_{t'' \in T} \mathbb{E}[y(t) - y(t') \mid z = t''] P(z = t'')$$

Sharp bounds for the unidentified expectations on the right hand side of (3.10) will yield sharp bounds on $\mathbb{E}[y(t) - y(t')]$. I proceed by finding the sharp identification region for an

arbitrary $y^i(t) - y^i(t')$ and every possible z^i . These can be averaged to find sharp bounds on $E[y(t) - y(t') | z]$ for all z . Define the following sets:

$$\begin{aligned}
\Lambda_1 &= \{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma_{t,t'}^{MOD} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t' < t < t'' \text{ and} \\
&\text{either } \exists \gamma \in \Gamma_{t,t'}^{MOD} \text{ s.t. } t'' \in S_\gamma \text{ or } \exists \gamma \in \Gamma_{t,t'}^{SPM}, \gamma' \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma \cap S_{\gamma'}\} \\
\Lambda_2 &= \{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t' < t < t'', \\
&\nexists \gamma \in \Gamma_{t,t'}^{MOD} \cup \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma\} \\
\Lambda_3 &= \{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t' < t < t'', \\
&\nexists \gamma \in \Gamma_{t,t'}^{MOD} \cup \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma\} \\
(3.11) \quad \Lambda_4 &= \{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{MOD} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t'' < t' < t \text{ and} \\
&\text{either } \exists \gamma \in \Gamma_{t,t'}^{MOD} \text{ s.t. } t'' \in S_\gamma \text{ or } \exists \gamma \in \Gamma_{t,t'}^{SPM}, \gamma' \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma \cap S_{\gamma'}\} \\
\Lambda_5 &= \{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t'' < t' < t, \\
&\nexists \gamma \in \Gamma_{t,t'}^{MOD} \cup \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma\} \\
\Lambda_6 &= \{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma\} \cup \{t'' \mid t'' < t' < t, \\
&\nexists \gamma \in \Gamma_{t,t'}^{MOD} \cup \Gamma_{t,t'}^{SPM} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma_{t,t'}^{SBM} \text{ s.t. } t'' \in S_\gamma\} \\
\Lambda_7 &= \left(\bigcup_{j=1}^6 \Lambda_j \right)^c
\end{aligned}$$

The four distinct orderings presented in $\Lambda_1, \dots, \Lambda_6$ in (3.11) include every possible ordering that is compatible with the restriction that $t' < t$ and that also allows at least one of SPM or SBM to have some implications for identification. Each of $\Lambda_1, \dots, \Lambda_6$ is a union of two sets. When t'' is not comparable with exactly one of t, t' , there can be at most one four-point sublattice containing t, t' , and t'' , since the incomparable treatments define a unique meet and join. This simplifies the construction of the first set in each of these six two-set unions. The first set in Λ_1 isolates the t'' which belong to a sublattice containing t and t' where both

SPM and SBM hold and where t'' is not strictly larger or smaller than t' , so it must be the case that $t' \vee t'' = t$. Since both SPM and SBM hold on this sublattice, it follows that

$$y^i(t) - y^i(t') = y^i(t'') - y^i(t' \wedge t'')$$

so that, when $z^i = t''$, the bounds

$$y^i(t) - y^i(t') \in [y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}]$$

are sharp. A similar argument can be made for the first set in each of $\Lambda_2, \dots, \Lambda_6$, and these sets are mutually exclusive due to the particular combinations of order restrictions and γ memberships along with the fact that

$$\Gamma_{t,t'}^{MOD} \cap \Gamma_{t,t'}^{SPM} = \Gamma_{t,t'}^{MOD} \cap \Gamma_{t,t'}^{SBM} = \Gamma_{t,t'}^{SPM} \cap \Gamma_{t,t'}^{SBM} = \emptyset$$

by definition.

The construction of the second set in each of the six two-set unions $\Lambda_1, \dots, \Lambda_6$ is complicated by the fact that the orderings $t' < t < t''$ and $t'' < t' < t$ are compatible with multiple four-point sublattices containing t , t' , and t'' , since there may be multiple t''' such that $t \vee t''' = t''$ and $t \wedge t''' = t'$ (in the former case) and $t' \vee t''' = t$ and $t' \wedge t''' = t''$ (in the latter case). Each set is constructed to capture the t'' whose sublattice membership(s) yield the same implications for identification as the set it is paired with. The particular combinations of order restrictions and γ memberships imply that they are mutually exclusive.

The sets $\Lambda_1, \dots, \Lambda_6$ define every sublattice membership pattern for t and t' for which SPM and SBM may have any implications; this follows from proposition 1 and its straightforward extension to the case of SBM. The set Λ_7 contains those t'' such that either (1) any sublattice S_γ containing t , t' , and t'' must have t' as the bottom and t as the top, (2) t'' does not belong to any four-point sublattice containing t and t' , or (3) t'' obeys one of the orderings from $\Lambda_1, \dots, \Lambda_6$ but does not belong to any sublattice containing t and t' on which at least one of SPM and SBM hold.

The focus on four-point sublattices is without loss of generality, since the implications of assumptions SPM and SBM only appear on four-point sublattices. SPM and SBM have no implications on chains, so sublattices that are chains can be ignored. Restricting attention to elements of $\{S_\gamma\}_{\gamma \in \Gamma_{t,t'}} \subseteq \{S_\gamma\}_{\gamma \in \Gamma}$ is without loss of generality as well. This follows from the fact that SPM and SBM have no implications for potential outcomes under the maintained assumptions, and any implications for the treatment effect $y^i(t) - y^i(t')$ from another treatment effect which are mediated by a third treatment effect are realized directly on a sublattice containing the treatments from the first two treatment effects. To see this concretely, suppose that $S_\gamma = \{t', t, t'', t'''\}$ and $S_{\gamma'} = \{t'', t''', t'''', t''''\}$ where $t \parallel t'', t''' \parallel t''''$, $t' = t \wedge t'', t''' = t \vee t'', t'' = t''' \wedge t''''$, and $t'''' = t''' \vee t''''$. Suppose that SPM holds on both S_γ and $S_{\gamma'}$. This implies

$$\begin{aligned} y^i(t) - y^i(t') &\leq y^i(t''') - y^i(t'') \leq y^i(t'') - y^i(t''') \\ &\implies y^i(t) - y^i(t') \leq y^i(t'') - y^i(t''') \end{aligned}$$

I show that $\{t', t, t''', t''''\} \in \{S_\gamma\}_{\gamma \in \Gamma_{t,t'}}^{SPM}$; this follows directly from lemma 1 and the definition of $\Gamma_{t',t}^{SPM}$:

Lemma 1. *Assume that $t \parallel t'', t''' \parallel t''''$, $t' = t \wedge t'', t''' = t \vee t'', t'' = t''' \wedge t''''$, and $t'''' = t''' \vee t''''$. Then, $t''' \wedge t = t'$ and $t'''' \vee t = t''''$.*

Proof. See appendix. □

A similar argument applies for SBM.

The sets defined in (3.11) along with the arguments of proposition 1 yield the following sharp identification regions for $y^i(t) - y^i(t')$ and each possible z^i :

$$(3.12) \quad y^i(t) - y^i(t') \in \begin{cases} [y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i \in \{t\} \cup \Lambda_1 \\ [\underline{K} - \overline{K}, y^i(z^i) - \underline{K}] & \text{if } z^i \in \Lambda_2 \\ [y^i(z^i) - \overline{K}, \overline{K} - \underline{K}] & \text{if } z^i \in \Lambda_3 \\ [\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)] & \text{if } z^i \in \{t'\} \cup \Lambda_4 \\ [\underline{K} - y^i(z^i), \overline{K} - \underline{K}] & \text{if } z^i \in \Lambda_5 \\ [\underline{K} - \overline{K}, \overline{K} - y^i(z^i)] & \text{if } z^i \in \Lambda_6 \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z^i \in \Lambda_7 \end{cases}$$

Since the sets $\{t\}, \{t'\}, \Lambda_1, \dots, \Lambda_7$ are mutually exclusive and exhaustive, averaging the bounds in (3.12) across i yields sharp bounds on $E[y(t) - y(t')]$ via (3.10). These sharp bounds are given in (3.9).

□

Proposition 2 generalizes proposition 1 by allowing for a much richer set of treatments. The treatment may have any finite number of dimensions, and each may be binary or multivalued. Some dimensions of the treatment may be complements while others are substitutes; the result allows for arbitrary combinations of SPM and SBM as appropriate. The complexity of the result is due to two factors. First, the treatment pair t, t' may belong to multiple sublattices. Second, the position of the treatment pair within a lattice, i.e., whether it includes the top and/or bottom of the sublattice, differs across sublattices. The position of the treatment pair within a sublattice combined with the assumptions that hold on the sublattice determine whether the upper and/or lower bound (or neither) are improved.

A number of the assumptions made in proposition 2 are primarily for ease of exposition and interpretation and do not limit the generality of the result. For example, the assumption that each S_γ has a cardinality of four is without loss of generality, since the implications of

assumptions SPM and SBM only appear on four-point sublattices. Similarly, no generality is sacrificed by excluding sublattices that are chains, as SPM and SBM have no implications on chains.

I have focused on bounding expectations of treatment effects using only supermodularity and submodularity assumptions, but in practical applications these will often be paired with other monotonicity and instrumental variable assumptions. Deriving sharp bounds under combinations of assumptions is nontrivial. Applying results from section 4 to bound $E[y(t) - y(t') \mid z = t'']$ for each $t, t', t'' \in T$ before applying proposition 2 will yield bounds that contain the true value but are not necessarily sharp. However, these bounds may be much simpler to compute than the sharp bounds (which remain an open problem).

4. INSTRUMENTAL VARIABLES

Traditional instrumental variable (IV) analysis of treatment response relies on the existence of a variable that is correlated with the treatment variable of interest but is mean-independent or independent of the distribution of response functions. Whether or not such independence assumptions are justified in a particular context is often the subject of vigorous debate. This has motivated researchers to find weaker and more credible forms of these assumptions that still retain some identification power. A leading example is the notion of a monotone instrumental variable (Manski and Pepper 2000, 2009):

Assumption MIV. x_k is a *monotone instrumental variable* if, for all $t, t' \in T$ and all x_{-k} ,

$$x_k \leq x'_k \implies E[y(t) \mid z = t', x = (x_k, x_{-k})] \leq E[y(t) \mid z = t', x = (x'_k, x_{-k})]$$

Manski and Pepper motivated MIV by considering the problem of determining the returns to schooling. Average wages should be weakly increasing with observable measures of ability (such as test scores or realized years of schooling), so such measures can be used as MIVs but not IVs. Giustinelli (2011) analyzes the returns to education in Italy using a similar monotonicity restriction on the quantile function.

Assumption MIV can be generalized to allow x_k to be partially ordered. Manski and Pepper then refer to x_k as a semi-monotone instrumental variable (SMIV). Another special case of MIV occurs when the realized treatment z is itself an MIV; Manski and Pepper refer to this as the monotone treatment selection (MTS) assumption. MIV and its generalizations impose restrictions on functionals of potential outcome distributions. Restrictions can also be imposed directly on functionals of treatment effect distributions:

Assumption SPMIV (Supermodular instrumental variable). x_k is a *supermodular instrumental variable* for $E[y(t) - y(t') \mid x_k, x_{-k}]$ with $t' \leq t$ if

$$(4.1) \quad x_k \leq x'_k \implies E[y(t) - y(t') \mid x_k, x_{-k}] \leq E[y(t) - y(t') \mid x'_k, x_{-k}]$$

for all x_{-k} .¹³

SPMIV is an alternative formulation of complementarity where treatment effects vary monotonically (on average) with an observed covariate x_k .¹⁴ An advantage of these assumptions is that evidence for their validity may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. This evidence can motivate the application of these assumptions in other contexts where similar identification strategies are not available. This contrasts with traditional IV assumptions, which tend to be highly context-specific.

The Djebbari and Smith (2008) study of the heterogeneous impacts of the PROGRESA conditional cash transfer program provides some examples of potential SPMIVs. PROGRESA provided payments to households conditional on regular school attendance by the household's children as well as visits to health centers. Djebbari and Smith find that the impact of this program on per capita consumption is substantially larger for poorer households and households in more "marginal" villages, i.e., villages with greater rates of illiteracy, more limited infrastructure, and a greater dependence on agricultural activities. Evaluations of

¹³The weak inequality in (4.1) can be reversed, in which case x_k would be a *submodular instrumental variable*. If the inequality is replaced with equality, x_k becomes a *modular instrumental variable*.

¹⁴The SPM/SPMIV distinction is analogous to the MTR/MIV distinction discussed in Manski and Pepper (2009).

cash transfer programs in other contexts could make use of this information by using household poverty or village marginality as SPMIVs.

Further examples are provided by the Bitler et al. (2014) study of the impact of the Connecticut Jobs First experiment. This program substantially lowered the marginal tax rate on earnings below the poverty line for families on relief, relative to the existing Aid to Families with Dependent Children (AFDC) program. In the Jobs First program, the entire benefit package is terminated once earnings rise above the poverty line; this is in contrast to the AFDC, where benefits decline linearly with earnings. Labor supply theory clearly suggests that the impact of this alternative budget scheme should boost earnings and employment much more for those who were previously out of work or whose earnings left them far below the poverty line. These hypotheses are strongly borne out by the data, suggesting that measures of pre-program earnings and employment could serve as SPMIVs in studies of similar programs which are not implemented experimentally.

For the remainder of this section, let $\underline{B}(t, x)$ and $\overline{B}(t, x)$ be defined as

$$\underline{B}(t, x) = E[y(t) | z = t, x] P(z = t | x) + \underline{K} P(z \neq t | x) \quad \forall t \in T, x \in X$$

and

$$\overline{B}(t, x) = E[y(t) | z = t, x] P(z = t | x) + \overline{K} P(z \neq t | x) \quad \forall t \in T, x \in X$$

The following bounds can be derived using SPMIV:

Proposition 3. *Assume that x_k is an SPMIV for $E[y(t) - y(t') | x_k, x_{-k}]$ with $t, t' \in T$.*

Then, the bounds

$$(4.2) \quad \begin{aligned} & \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} \\ & \leq E[y(t) - y(t') | x_k, x_{-k}] \leq \\ & \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} \end{aligned}$$

are sharp.

As is the case for bounds derived under IV or MIV assumptions, inference is complicated by the sup and inf operators in equation (4.2) (Manski and Pepper 2009). Analog estimators of the bounds in (4.2) are consistent but biased in finite samples; the estimated bounds will generally be too narrow. Fortunately, the methods developed by Chernozhukov, Lee and Rosen (2013) can be applied to find bias-corrected estimates and associated confidence intervals. Chernozhukov et al. discuss in detail the special cases of estimating nonparametric bounds using instrumental variables and MIVs; the bounds in (4.2) are essentially identical for the purposes of estimation, so their results can be applied directly to my estimation problem. The theoretical extension allowing for multiple SPMIVs is straightforward, and presents no novel estimation challenges besides those associated with high-dimensional nonparametric conditioning.

Returning to assumption SPMIV: If the second inequality in (4.1) is reversed, x_k becomes a *submodular instrumental variable*. If x_k is a supermodular and submodular instrumental variable, i.e., average treatment effects are constant across different values of x_k , then x_k is a *modular instrumental variable*. While this may seem like a strong assumption, it is routinely employed in applied work that assumes both exogeneity of the treatment and no interactions.

SPMIVs may also improve the bounds on functionals of potential outcome distributions, as the following proposition illustrates:

Proposition 4. *Assume that x_k is an SPMIV for $E[y(t) - y(t') \mid x_k, x_{-k}]$ with $t, t' \in T$. Then, the bounds*

$$\begin{aligned}
 (4.3) \quad & \max \left\{ \underline{B}(t, x), \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} + \underline{B}(t', x) \right\} \\
 & \leq E[y(t) \mid x_k, x_{-k}] \leq \\
 & \min \left\{ \overline{B}(t, x), \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} + \overline{B}(t', x) \right\}
 \end{aligned}$$

and

$$\begin{aligned}
(4.4) \quad & \max \left\{ \underline{B}(t', x), \underline{B}(t, x) - \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} \right\} \\
& \leq E[y(t') \mid x_k, x_{-k}] \leq \\
& \min \left\{ \overline{B}(t', x), \overline{B}(t, x) - \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} \right\}
\end{aligned}$$

are sharp.

As in the case of proposition 3, analog estimators of the bounds in (4.3) and (4.4) are consistent but biased in finite samples; the Chernozhukov et al. approach can be applied here as well.

5. INDEPENDENCE

Independence assumptions have been used to operationalize the belief that individuals' realized treatments are unrelated to any individual characteristics which may influence responses. This should be the case, for example, in a randomized controlled trial. I show how statistical independence can be combined with shape restrictions and instrumental variables assumptions to narrow the bounds on entire treatment effect distributions.

The familiar assumption of statistical independence of treatments and response functions is defined in my notation as follows:

Assumption SI (Statistical independence). Potential outcomes are *statistically independent* of realized treatments if

$$P(y(t) \mid z) = P(y(t)) \quad \forall t \in T$$

Assumption SI implies that the marginal distribution of $y(t)$, denoted F_t , is point identified for all $t \in T$ such that $P(z = t) > 0$. However, the distribution of $y(t) - y(t')$, whose cumulative distribution function is denoted by $F_{t,t'}$, is only partially identified. Makarov (1982) was the first to derive pointwise sharp bounds on the distribution of the sum of two random variables with fixed marginal distributions. Frank, Nelsen and Schweizer (1987) derived these

bounds in a simpler manner and extended them to allow for other operations such as differences and products as well as more than two variables. However, as Kreinovich and Ferson (2006) show, these bounds are not sharp in the case of more than two variables. The following result, taken from Theorem 2 of Williamson and Downs (1990), gives the pointwise sharp bounds on the distribution $F_{t,t'}$ for any $t, t' \in T^R$:

$$\begin{aligned}
(5.1) \quad \underline{F}_{t,t'}(w) &= \sup_{u+v=w} \{ \max \{ F_t(u) - F_{t'}(-v), 0 \} \} \\
&\leq F_{t,t'}(w) \leq \\
&1 + \inf_{u+v=w} \{ \min \{ F_t(u) - F_{t'}(-v), 0 \} \} = \overline{F}_{t,t'}(w)
\end{aligned}$$

Fan and Park (2010) discuss consistent nonparametric estimation of these bounds.

SI can be combined with SPM to refine (5.1), as the following result shows:

Proposition 5. *Assume that SI holds and that $T = \{t \wedge t', t, t', t \vee t'\}$ with $t \wedge t' < t, t' < t \vee t'$. Assume that SPM holds on T . Then, the bounds*

$$(5.2) \quad \underline{F}_{t \vee t', t}(w) \leq F_{t \vee t', t}(w) \leq \min \{ \overline{F}_{t \vee t', t}(w), \overline{F}_{t', t \wedge t'}(w) \}$$

and

$$\max \{ \underline{F}_{t, t \wedge t'}(w), \underline{F}_{t \vee t', t'}(w) \} \leq F_{t, t \wedge t'}(w) \leq \overline{F}_{t, t \wedge t'}(w)$$

are sharp, where $\underline{F}, \overline{F}$ are defined as in (5.1).

Similar results can be derived for SBM, and these results can be used to obtain narrower bounds on functionals of treatment effect distributions.¹⁵ These shape restrictions could be justified by theoretical arguments; alternatively, since average treatment effects are point-identified in this context, the supermodularity or submodularity of average effects could be used to provide some justification for stronger structural assumptions. Extending these

¹⁵However, see Firpo and Ridder (2010) for a discussion of pointwise vs. uniform sharpness and the implications for deriving sharp bounds on functionals of the distribution of treatment effects.

results to general lattices is problematic due to the fact that sharp bounds on the distribution function of a sum of more than two variables are an open question. Nonetheless, it is straightforward to collect all possible stochastic dominance relations implied by the maintained assumptions, and bounds which contain the true value (but are not necessarily sharp) can be obtained in a manner similar to that in proposition 5. Such bounds may be useful in policy evaluation.

A reformulation of the SPMIV assumption can also be applied in this setting. SPMIV itself is unhelpful, since conditional average treatment effects are point-identified. However, if the distribution of treatment effects conditional on x is thought to obey a stochastic dominance relationship in one or more covariates, this can be used to derive improved bounds. The formal statement of the assumption is as follows:

Assumption Q-SPMIV (Quantile supermodular instrumental variable). x_k is a *quantile supermodular instrumental variable* for $y(t) - y(t')$ if

$$x_k \leq x'_k \implies F_{t,t'}(w \mid x_k, x_{-k}) \geq F_{t,t'}(w \mid x'_k, x_{-k})$$

for all x_{-k} .

The following proposition computes the bounds derived under this assumption:

Proposition 6. *Assume that SI holds. Assume that x_k is a Q-SPMIV for $y(t) - y(t')$ with $t, t' \in T$. Then, the bounds*

$$\sup_{x_k \leq x'_k} \{F_{t,t'}(w \mid x'_k, x_{-k})\} \leq F_{t,t'}(w \mid x_k, x_{-k}) \leq \inf_{x'_k \leq x_k} \{\overline{F}_{t,t'}(w \mid x'_k, x_{-k})\}$$

are sharp.

Again, since conditional average treatment effects are point-identified, they can provide some evidence to support the validity of the stronger Q-SPMIV assumption. The improved bounds on $F_{t,t'}$ derived using this result can be combined with SPM or SBM to yield even stronger bounds.

6. EMPIRICAL ILLUSTRATION

To illustrate the use of the identification results developed in this paper, I reanalyze data from Shertzer et al. (2014). That study examines the extent to which Chicago’s first zoning ordinance, passed in 1923, influenced the evolution of the spatial distribution of commercial, industrial, and residential activity in the city. That study found evidence of substantial treatment effect heterogeneity, which motivates the use of SBM and SPMIV assumptions in the analysis below.

Chicago’s 1923 zoning ordinance regulated land by restricting uses and density; for details on the ordinance, consult Shertzer et al. (2014). Here, I bound the effects of 1923 commercial zoning on the probability that a city block will contain any commercial activity in 2005, focusing on the outlying (largely residential) portions of the city that were zoned into the two lowest density categories. As discussed in section 3, zoning is a multidimensional policy and the long-run effect of commercial zoning likely varies substantially with the associated density restrictions. Since both use and density zoning are endogenous policy variables, quantifying the heterogeneous effects of commercial zoning with respect to density requires a multidimensional treatment variable; simply conditioning on assigned density zoning would not yield correct estimates of how the commercial zoning effect varies with density zoning.

Formally, the outcome variable $y_i(\cdot)$ is an indicator equal to 1 iff city block i contains any commercial activity in 2005. y_i is a function of a treatment $t \in T = \{0, 1\} \times \{1, 2\}$. The first dimension of t is equal to 1 if the block received any commercial zoning in 1923 and 0 otherwise. The second dimension of t is equal to 1 if the block was zoned for the lowest density development (3 or fewer stories) and 2 if it was zoned for higher density development (8–10 stories).

Areas zoned for lower densities will be more residential in character and contain a larger proportion of single-family homes (Shertzer et al. 2014). It is well documented that residential property owners (especially single-family homeowners) generally oppose the encroachment of commercial uses and have substantial power to block such development (Fischel 2001). It is likely that the early establishment of commercial activity through zoning will be

a more important determinant of future commercial land use in areas also zoned for lower densities. This assumption is also consistent with previous literature showing that mixed use areas are more likely to see conversion to completely non-residential use than strictly residential use (McMillen and McDonald 1991). This motivates the assumption that y exhibits SBM on T .

One may also expect commercial zoning to have more persistent effects when it does not conflict with the existing land use pattern. The data identifies blocks which had commercial activity prior to the introduction of zoning; an indicator for the presence of pre-zoning commercial activity is a natural SPMIV.

Assumptions:	Bounds on:	
	$E[y(1, 1) - y(0, 1)]$	$E[y(1, 2) - y(0, 2)]$
None	$[-0.686, 0.934]$	$[-0.573, 0.807]$
SBM	$[-0.259, 0.934]$	$[-0.573, 0.741]$
SPMIV	$[-0.532, 0.934]$	$[-0.573, 0.747]$
SBM & SPMIV	$[-0.231, 0.934]$	$[-0.573, 0.512]$
Observations	8,572	8,572

TABLE 1. The outcome variable y is an indicator for the presence of commercial activity on the block in 2005. The first dimension of the treatment is an indicator that equals one iff the block received any commercial zoning in 1923. The second dimension of the treatment indicates the density level the block was zoned for in 1923.

Table 1 shows a series of bounds computed under different assumptions. It is noteworthy that the bounds under the combination of SBM and SPMIV are not simply the intersection of the bounds computed under these assumptions separately. This illustrates the fact that the shape restrictions I introduce can magnify the identifying power of other assumptions. While the sign of the treatment effect is not identified using only these assumptions, there is up to a 28% reduction in the width of the bounds.

7. CONCLUSION

In this paper, I contribute to the literature on the partial identification of treatment effects by developing and applying assumptions that formalize the notion of complementarity.

I examine the identification power of these assumptions and discuss how they can be justified. The supermodularity and submodularity assumptions I propose can be used to narrow bounds on treatment effects in studies of policy complementarity, which have traditionally been stymied by a lack of pseudo-experimental variation in multiple policies simultaneously. In proposition 1, I show how these shape restrictions can improve bounds on average treatment effects in the simple case of a two-dimensional binary treatment. Proposition 2 extends this result to a more general treatment set with an arbitrary finite number of (possibly multivalued) treatments and the possibility of complex combinations of supermodularity and submodularity.

Complementarity may also stem from differential treatment response among subpopulations defined by observed covariates. Subgroup heterogeneity in treatment effects is an increasingly widely recognized phenomenon, and can often be motivated directly from economic theory (see, e.g., Bitler et al. (2014)). Propositions 3 and 4 show how qualitative information about treatment effect heterogeneity embodied in supermodular instrumental variables can be used to improve bounds on average treatment effects and average potential outcomes. Supermodular instrumental variables can be used in studies with one or many treatments, making them a versatile and potentially powerful addition to the arsenal of applied econometricians.

The assumptions I propose can be useful in the experimental context as well. Proposition 5 shows how supermodularity can be combined with an assumption of statistical independence between assigned treatments and responses to yield improved bounds on the cumulative distribution function of a treatment effect. These results can be applied to the evaluation of outcomes in complex (multi-treatment) randomized controlled trials, which are increasingly prevalent in many fields, including development economics. Since average treatment effects are point-identified in this context, one can determine if average responses exhibit supermodularity or submodularity. This can provide evidence that individual response functions are supermodular or submodular. Similarly, the behavior of (point-identified) conditional

average treatment effects can motivate the use of a quantile supermodular instrumental variable; in proposition 6, I show how this assumption can strengthen the bounds on the CDF of a treatment effect distribution.

Bounds derived under the assumptions I propose here are of interest only to the extent that such assumptions are considered credible. Where might evidence for their validity come from? Arguments for policy complementarity may be provided by economic theory, as in Lalive et al. (2006), or they may come from multi-treatment randomized controlled trials. Evidence on subgroup heterogeneity in treatment effects may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. In such studies, conditional average treatment effects are point-identified, so the validity of the assumptions I propose can be established. This can motivate their use in other contexts where similar identification strategies are not available. This distinguishes supermodular IV assumptions from traditional IV assumptions, since the latter tend to be context-specific.

The empirical illustration in section 6 employs assumptions SBM and SPMIV to study the impact of historical zoning on the evolution of land use in Chicago. Of particular interest is the fact that the bounds computed under both SBM and SPMIV are substantially narrower than the intersection of the bounds computed under each assumption separately. This demonstrates the general fact that assumptions SPM and SBM can magnify the identification power of other assumptions.

8. APPENDIX

Proof of lemma 1. I first show that $t'''' \wedge t = t'$:

$$\begin{aligned}
& t'''' \wedge t''' = t'' \\
\implies & (t'''' \wedge t''') \wedge t = t'' \wedge t \\
\implies & t'''' \wedge (t''' \wedge t) = t' \\
\implies & t'''' \wedge t = t'
\end{aligned}$$

where the last implication follows from the fact that $t \leq t'''$. Now, I show that $t''' \vee t = t''''$:

$$\begin{aligned}
t''' \vee t''' &= t'''' \\
\implies (t \vee t'') \vee t''' &= t'''' \\
\implies t \vee (t'' \vee t''') &= t'''' \\
\implies t \vee t''' &= t''''
\end{aligned}$$

where the last implication follows from the fact that $t'' \leq t'''$. □

Proof of proposition 3. In the absence of other assumptions, the bounds

$$\underline{B}(t, x) \leq E[y(t) \mid x] \leq \overline{B}(t, x)$$

and

$$\underline{B}(t', x) \leq E[y(t') \mid x] \leq \overline{B}(t', x)$$

and thus

$$\underline{B}(t, x) - \overline{B}(t', x) \leq E[y(t) \mid x] - E[y(t') \mid x] \leq \overline{B}(t, x) - \underline{B}(t, x)$$

are sharp for all $x \in X$. The assumption that x_k is an SPMIV for $E[y(t) - y(t') \mid x_k, x_{-k}]$ implies that

$$\underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \leq E[y(t) - y(t') \mid x_k, x_{-k}]$$

for all $x'_k \leq x_k$ and

$$E[y(t) - y(t') \mid x_k, x_{-k}] \leq \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k})$$

for all $x_k \leq x'_k$. The result follows. □

Proof of proposition 4. Proposition 3 implies that the bounds

$$\sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \}$$

$$\leq E[y(t) - y(t') \mid x_k, x_{-k}] \leq$$

$$\inf_{x_k \leq x'_k} \{\overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k})\}$$

are sharp. Thus, $E[y(t) \mid x]$ and $E[y(t') \mid x]$ must simultaneously satisfy the no-assumption bounds

$$(8.1) \quad \underline{B}(t, x) \leq E[y(t) \mid x] \leq \overline{B}(t, x)$$

and

$$(8.2) \quad \underline{B}(t', x) \leq E[y(t') \mid x] \leq \overline{B}(t', x)$$

as well as

$$(8.3) \quad \sup_{x'_k \leq x_k} \{\underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k})\} + E[y(t') \mid x]$$

$$\leq E[y(t) \mid x] \leq$$

$$\inf_{x_k \leq x'_k} \{\overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k})\} + E[y(t') \mid x]$$

and

$$(8.4) \quad E[y(t) \mid x] - \inf_{x_k \leq x'_k} \{\overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k})\}$$

$$\leq E[y(t') \mid x] \leq$$

$$E[y(t) \mid x] - \sup_{x'_k \leq x_k} \{\underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k})\}$$

From (8.1)–(8.4), it is clear that

$$\max \left\{ \underline{B}(t, x), \sup_{x'_k \leq x_k} \{\underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k})\} + \underline{B}(t', x) \right\}$$

$$\leq E[y(t) \mid x_k, x_{-k}] \leq$$

$$\min \left\{ \overline{B}(t, x), \inf_{x_k \leq x'_k} \{\overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k})\} + \overline{B}(t', x) \right\}$$

and

$$\begin{aligned} & \max \left\{ \underline{B}(t', x), \underline{B}(t, x) - \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} \right\} \\ & \leq \mathbb{E}[y(t') \mid x_k, x_{-k}] \leq \\ & \min \left\{ \overline{B}(t', x), \overline{B}(t, x) - \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} \right\} \end{aligned}$$

must hold. I show that these bounds are feasible, i.e., consistent with (4.2), whence it follows that they are sharp. Consider the following events:

$$\begin{aligned} (8.5) \quad & \max \left\{ \underline{B}(t, x), \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} + \underline{B}(t', x) \right\} \\ & = \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} + \underline{B}(t', x) > \underline{B}(t, x) \end{aligned}$$

$$\begin{aligned} (8.6) \quad & \min \left\{ \overline{B}(t, x), \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} + \overline{B}(t', x) \right\} \\ & = \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} + \overline{B}(t', x) < \overline{B}(t, x) \end{aligned}$$

$$\begin{aligned} (8.7) \quad & \max \left\{ \underline{B}(t', x), \underline{B}(t, x) - \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} \right\} \\ & = \underline{B}(t, x) - \inf_{x_k \leq x'_k} \{ \overline{B}(t, x'_k, x_{-k}) - \underline{B}(t', x'_k, x_{-k}) \} > \underline{B}(t', x) \end{aligned}$$

$$\begin{aligned} (8.8) \quad & \min \left\{ \overline{B}(t', x), \overline{B}(t, x) - \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} \right\} \\ & = \overline{B}(t, x) - \sup_{x'_k \leq x_k} \{ \underline{B}(t, x'_k, x_{-k}) - \overline{B}(t', x'_k, x_{-k}) \} < \overline{B}(t', x) \end{aligned}$$

It is easy to show that (8.5) $\implies \neg(8.7)$; thus, the lower bounds in (4.3) and (4.4) are consistent with (4.2). Similarly, (8.6) $\implies \neg(8.8)$, and so the upper bounds in (4.3) and (4.4) are consistent with (4.2). \square

Proof of proposition 5. For a lattice $T = \{t, t', t \vee t', t \wedge t'\}$ which is not a chain, SPM implies the following inequalities:

$$y^i(t') - y^i(t \wedge t') \leq y^i(t \vee t') - y^i(t)$$

$$y^i(t) - y^i(t \wedge t') \leq y^i(t \vee t') - y^i(t')$$

$$y^i(t) + y^i(t') - 2y^i(t \wedge t') \leq y^i(t \vee t') - y^i(t \wedge t') \leq 2y^i(t \vee t') - y^i(t) - y^i(t')$$

Since these inequalities hold for all $i \in I$, they imply the following first-order stochastic dominance relationships:

$$(8.9) \quad F_{t \vee t', t}(w) \leq F_{t', t \wedge t'}(w)$$

$$(8.10) \quad F_{t \vee t', t'}(w) \leq F_{t, t \wedge t'}(w)$$

$$F_{t \vee t', t, t \vee t', t'}(w) \leq F_{t \vee t', t \wedge t'}(w)$$

$$F_{t \vee t', t \wedge t'}(w) \leq F_{t, t \wedge t', t', t \wedge t'}(w)$$

for all $w \in \mathbb{R}$. Here, $F_{t \vee t', t, t \vee t', t'}$ is the cdf of $2y(t \vee t') - y(t) - y(t')$ and $F_{t, t \wedge t', t', t \wedge t'}$ is the cdf of $y(t) + y(t') - 2y(t \wedge t')$. Pointwise sharp bounds on $F_{t \vee t', t}$, $F_{t \vee t', t'}$, $F_{t', t \wedge t'}$, $F_{t, t \wedge t'}$, and $F_{t \vee t', t \wedge t'}$ in the absence of SPM are given by (5.1). Combining SPM with the inequalities (8.9) and (8.10) yields the results. □

Proof of proposition 6. Trivial. □

REFERENCES

- Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business & Economic Statistics*, **26**(3), pp. 261–283.
- Beresteanu, A. (2005). Nonparametric analysis of cost complementarities in the telecommunications industry. *RAND Journal of Economics*, **36**(4), pp. 870–889.
- (2007). Nonparametric estimation of regression functions under restrictions on partial derivatives. *Unpublished manuscript*.
- Bhattacharya, J., A.M. Shaikh, and E. Vytlacil (2008). Treatment effect bounds under monotonicity assumptions: An application to Swan–Ganz catheterization. *American Economic Review*, **98**(2), pp. 351–356.

- Bhattacharya, Jay, Azeem M. Shaikh, and Edward Vytlacil (2012). Treatment effect bounds: An application to Swan–Ganz catheterization. *Journal of Econometrics*, **168**(2), pp. 223–243.
- Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, **96**(4), pp. 988–1012.
- (2008). Distributional impacts of the self-sufficiency project. *Journal of Public Economics*, **92**(3), pp. 748–765.
- (2014). Can variation in subgroups’ average treatment effects explain treatment effect heterogeneity? Evidence from a social experiment. *NBER Working Paper 20142*.
- Cattaneo, Matias D. (2010). Multi-valued treatment effects. In Steven N. Durlauf and Lawrence E. Blume, eds., *New Palgrave Dictionary of Economics*, Palgrave Macmillan.
- Chernozhukov, V., S. Lee, and A.M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica*, **81**(2), pp. 667–737.
- Cox, D.R. (1958). *Planning of experiments*. Wiley.
- Djebbari, H. and J. Smith (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, **145**(1), pp. 64–80.
- Eeckhout, J. and P. Kircher (2011). Identifying sorting – In theory. *Review of Economic Studies*, **78**(3), pp. 872–906.
- Fan, Y. and S.S. Park (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, **26**(3), p. 931.
- Feller, A. and C.C. Holmes (2009). Beyond topline: Heterogeneous treatment effects in randomized experiments. *Unpublished manuscript*.
- Firpo, S.P. and G. Ridder (2010). Bounds on functionals of the distribution treatment effects. *Unpublished manuscript*.
- Fischel, William A. (2001). *The homevoter hypothesis*. Harvard University Press.
- Frank, M.J., R.B. Nelsen, and B. Schweizer (1987). Best-possible bounds for the distribution of a sum – A problem of Kolmogorov. *Probability Theory and Related Fields*, **74**(2), pp. 199–211.
- Gelman, A. (2013). The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective. *Unpublished manuscript*.
- Giustinelli, P. (2011). Non-parametric bounds on quantiles under monotonicity assumptions: With an application to the Italian education returns. *Journal of Applied Econometrics*, **26**(5), pp. 783–824.
- Graham, Bryan S., Guido W. Imbens, and Geert Ridder (2014). Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics*, **5**(1), pp. 29–66.
- Gundersen, Craig, Brent Kreider, and John V. Pepper (2012). The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. *Journal of Econometrics*, **166**(1), pp. 79–91.
- Heckman, J.J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, **64**(4), pp. 487–535.
- Kreider, Brent and Steven C. Hill (2009). Partially identifying treatment effects with an application to covering the uninsured. *Journal of Human Resources*, **44**(2), pp. 409–449.

- Kreider, Brent and John V. Pepper (2007). Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, **102**(478), pp. 432–441.
- Kreinovich, V. and S. Ferson (2006). Computing best–possible bounds for the distribution of a sum of several variables is NP–hard. *International Journal of Approximate Reasoning*, **41**(3), pp. 331–342.
- Lalive, Rafael, Jan Van Ours, and Josef Zweimüller (2006). How changes in financial incentives affect the duration of unemployment. *Review of Economic Studies*, **73**(4), pp. 1009–1038.
- Lazzati, Natalia (2014). Treatment response with social interactions: Partial identification via monotone comparative statics. *Unpublished manuscript*.
- Makarov, G.D. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, **26**(4), pp. 803–806.
- Manski, C.F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, **24**(3), pp. 343–360.
- (1990). Nonparametric bounds on treatment effects. *American Economic Review*, **80**(2), pp. 319–323.
- (1997). Monotone treatment response. *Econometrica*, **65**(6), pp. 1311–1334.
- (2003). *Partial identification of probability distributions*. Springer.
- (2013). Identification of treatment response with social interactions. *Econometrics Journal*, **16**(1), pp. S1–S23.
- Manski, Charles F. and Daniel S. Nagin (1998). Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological methodology*, **28**(1), pp. 99–137.
- Manski, C.F. and J.V. Pepper (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, **68**(4), pp. 997–1010.
- (2009). More on monotone instrumental variables. *Econometrics Journal*, **12**, pp. S200–S216.
- McMillen, D.P. and J.F. McDonald (1991). A Markov chain model of zoning change. *Journal of Urban Economics*, **30**(2), pp. 257–270.
- Molinari, F. and A.M. Rosen (2008). The identification power of equilibrium in games: The supermodular case. *Journal of Business & Economic Statistics*, **26**(3), pp. 297–302.
- Neumark, David and William Wascher (2011). Does a higher minimum wage enhance the effectiveness of the Earned Income Tax Credit? *Industrial and Labor Relations Review*, **64**(4), pp. 712–746.
- Pepper, John V. (2000). The intergenerational transmission of welfare receipt: A nonparametric bounds analysis. *Review of Economics and Statistics*, **82**(3), pp. 472–488.
- Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, **6**(1), pp. 34–58.
- Shertzer, A., T. Twinam, and R.P. Walsh (2014). Zoning and urban persistence. *Unpublished manuscript*.
- Tsunao, Okumura and Emiko Usui (2014). Concave–monotone treatment response and monotone treatment selection: With an application to the returns to schooling. *Quantitative Economics*, **5**(1), pp. 175–194.

Williamson, R.C. and T. Downs (1990). Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4(2), pp. 89–158.